BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 8: e137717
doi: 10.3897/biss.8.137717

OPEN ACCESS

Conference Abstract

Completing the Completeness Measure: the MIDS Specification

Elspeth Margaret Haston[‡], Mathias Dillen[§], Sam Leeflang[|,¶], Cat Chapman[#]

‡ Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom
§ Meise Botanic Garden, Meise, Belgium
| Naturalis Biodiversity Center, Leiden, Netherlands
¶ Distributed System of Scientific Collections - DiSSCo, Leiden, Netherlands
# University of Florida, Gainesville, United States of America

Corresponding author: Elspeth Margaret Haston (e.haston@rbge.org.uk)

Received: 23 Sep 2024 | Published: 24 Sep 2024

Citation: Haston E, Dillen M, Leeflang S, Chapman C (2024) Completing the Completeness Measure: the MIDS Specification. Biodiversity Information Science and Standards 8: e137717.
https://doi.org/10.3897/biss.8.137717

Abstract

The need to rapidly digitise millions of specimens in natural history collections has seen the development of a staged approach for data capture by institutions that have been carrying out mass digitisation projects. The cost and difficulty of transcribing labels associated with specimens has necessitated an approach that prioritises parts of the data based on research and curatorial requirements as well as ease of transcription and methodology available. This approach means that there is currently a wide variation in the completeness of data capture and imaging of specimens within and between collections.

The Biodiversity Information Standards (TDWG) Minimum Information about a Digital Specimen (MIDS) specification has been designed to provide a framework for measuring and monitoring the completeness of the data transcribed from the specimen and the presence of an image or other media file. This framework enables a consistent and standardised approach to the calculation of the level of digitisation of specimens within a collection and, by extension/extrapolation, to the collection itself.

© Haston E et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Four levels of digitisation have been defined in the MIDS specification including a pre-digitisation level. These have been developed to correspond to practical requirements and existing large-scale digitisation programmes.

> **MIDS level 0 (Bare):** A bare or skeletal record making the association between an identifier of a physical specimen and its digital representation, allowing for unambiguous attachment of all other information.
>
> **MIDS level 1 (Basic):** A basic record of specimen information enabling basic discoverability as well as how the user is permitted to use the data.
>
> **MIDS level 2 (Regular):** A regular level of information including data that have been agreed over time as essential for most scientific purposes.
>
> **MIDS level 3 (Extended):** An extended level of information about a specimen including identifiers enabling connections to be made to other data present or known about the specimen.

The scope of MIDS does not directly include the quality of data entered. Instead, MIDS can be seen as a tool to support the continuing improvement of data access and quality. It provides guidance for prioritising the data to be captured as well as recommendations for data standards and mapping structures. Here we present the progress in the development of the MIDS specification and its implementation.

The MIDS elements for each of the four levels have now been defined and a draft version is available here. A machine-readable mapping to Darwin Core and ABCD has been compiled following the Simple Standard for Sharing Ontological Mappings (SSSOM). The differing data relevances and priorities for different disciplines have been recognised and the inclusion of required elements at each level is discipline specific (biology, palaeontology and geology). In addition, MIDS level scores have been increasingly implemented in existing specimen management systems, allowing an evaluation of the levels achieved in various contexts.

There are two implementations of MIDS calculators that have been developed and are available to test. One, developed at the Meise Botanic Garden, uses a zipped Global Biodiversity Information Facility (GBIF) annotated archive (a commonly used GBIF download format based on Darwin Core Archive) and is available here. The other, developed at The Natural History Museum London, interacts directly with the GBIF API and is available here. We encourage users to try these and provide feedback to the TDWG MIDS Task Group.

Keywords

digitisation, standard, specification, minimum information, identifiers

Presenting author

Elspeth Haston

Presented at

SPNHC-TDWG 2024

Acknowledgements

The authors wish to acknowledge the input and support of all the members of the TDWG MIDS Task Group.

Conflicts of interest

The authors have declared that no competing interests exist.